Consumer Data Access and Corporate Innovation: Evidence of Mobile Apps

Abstract

Our study examines how mobile app–enabled access to customer data influences corporate innovation. Findings reveal that post-app launch, firms see significant improvements in innovation intensity, quality, and efficiency. These effects are more pronounced in apps with larger user bases, those collecting user-linked data, and in data-driven digital innovations. Interestingly, compliance with privacy regulations does not significantly alter these innovation outcomes. Moreover, app launches lead to more user-centric innovations aimed at enhancing consumer experience. Collectively, these insights underscore the pivotal role of mobile apps in reshaping R&D strategies towards data-driven, user-focused solutions.

JEL classification: G30, O31, O33, D83

Keywords: Mobile Apps, Customer data, Corporate Innovation, Data Privacy

1. Introduction

Innovation is a process of integrating and leveraging existing information to create new knowledge (Aghion and Jaravel, 2015), and customers, as pivotal sources of knowledge, play a crucial role in this process by providing insights that help translate their needs into new products (e.g., Nambisan, 2002; Chatterji and Fabrizio, 2012; Saldanha, Mithas, and Krishnan, 2017). While firms have attempted to engage a small group of lead users in innovation activities, either by co-developing products with them or by enabling them to innovate independently (e.g., Von Hippel, 2005; Dahl et al., 2015; Schweisfurth, 2017),[1] they had limited ways of obtaining detailed information about their customers before the widespread use of digital technologies (Ferracuti et al., 2024). As a result, our understanding remains limited regarding how access to vast customer data, a crucial asset for firms (Bian, Ma, and Tang, 2021; Veldkamp, 2023), influences their innovation performance. In this study, we fill this void by focusing on the availability of customer data to firms via mobile apps.

A mobile app, short for mobile application, is software designed to run on smartphones, tablets, and other mobile devices. The widespread adoption of mobile apps has surged in recent years, and millions of apps are available across platforms like the Apple App Store, Google Play Store, and third-party Android markets, offering services that cater to nearly every aspect of daily life. The State of Mobile 2024 report reveals that in 2023, an average of 489,000 apps were downloaded every minute globally, while users spent an average of five hours per day on mobile devices.[2] The significant rise in mobile app usage has certainly caught the attention of businesses, prompting many to launch their own apps over the past decade (Wu, 2023). These apps offer firms a direct line to their customers, allowing for a seamless and continuous interaction. By leveraging these apps, firms are able to gather extensive customer data, gaining

[1] This phenomenon is known as "user innovation".
[2] Source: https://sensortower.com/state-of-mobile-2024.

insights into consumer preferences, behaviors, and trends (Ferracuti et al., 2024).

Customer data from mobile apps has three notable features: relevant, proprietary and exclusive, and cost-effective. Relevance refers to the idea that app data reflects genuine customer interactions and experiences with a firm's products or services, making them difficult to replace with data from other sources. In the meanwhile, customer data collected via apps is proprietary and exclusive, meaning they are owned, managed, and controlled solely by the firm, deeply embedded in its business processes and customer interactions, highly specific to its value proposition (Davenport and Redman, 2020), and inaccessible to external parties or competitors. Finally, app data is highly cost-effective, as it is generated passively through routine business activities, requires minimal incremental investment, and offers scalable, real-time, and accurate insights that support strategic decision-making.[3]

Based on these features, we expect that customer data from mobile apps stimulates corporate innovation by supporting both technological feasibility and commercial viability, two key determinants of innovation success (Chatterjee et al., 2023). In the digital era, innovation, especially digital-related innovation, often relies heavily on large volumes of data for model training (Lehrer et al., 2018; Beraja, Yang, and Yuchtman, 2023). The effectiveness of machine learning models, including deep learning and generative AI, depends on the quality, diversity, and scale of the data they are trained on. App data, being large-scale, real-time, and context-specific, enables firms to effectively train, test, and refine digital technologies, making technical solutions more reliable and responsive to real-world conditions.[4] At the same time, app data enhances commercial viability by offering insights into customer behavior,

[3] We provide an extensive discussion on these features in Section 2.

[4] For instance, Apple plans to use imagery and data collected from its Apple Maps Look Around feature to train AI models (Liszewski, 2025) and social media platforms (TikTok, Instagram, Facebook, Twitter/X) utilize user data from their apps to train AI models that enhance content recommendation algorithms, personalize user feeds, and improve ad targeting (Merchant, 2023). In Appendix 1, we provide three patent application examples showing that firms use app data as data elements in developing the self-learning control method and system for air conditioners, a heating equipment control method, and a user churn prediction method, respectively.

preferences, and unmet needs, which helps firms identify innovation opportunities, reduce market uncertainty, and tailor offerings to diverse user segments. Together, these capabilities allow firms to assess whether innovations are not only buildable but also desirable, ultimately improving innovation outcomes. Therefore, we hypothesize that the launching of mobile apps leads to enhanced corporate innovation performance.

Nonetheless, we also recognise that the link between data access and innovation is nuanced; having more customer data doesn't automatically foster greater innovation and can, in some cases, be counterproductive. The "streetlight effect" proposed by Hoelzemann et al. (2024) highlights the possibility that firms, influenced by readily available data via apps, may concentrate on seemingly promising but ultimately suboptimal projects, thereby narrowing their exploration and potentially suppressing innovation. Moreover, with access to extensive customer data, firms often achieve rapid market expansion through precision marketing and differentiated pricing strategies. This reliance on data-driven approaches can diminish the incentive for innovation. In addition, app data could be biased towards the most active users. In this case, the abundance of user data does not inherently lead to higher-quality innovation due to sampling bias (Cao et al., 2024).

To test our hypothesis, we analyze a comprehensive dataset on mobile apps and corporate innovation in China, where apps are downloaded the most around the world, encompassing 24,661 firm-year observations from 2007 to 2021. Given that the common practice of applying the log transformation to patent or citation counts may introduce bias (Cohn, Liu, and Wardlaw, 2022; Chen and Roth, 2024), we employ a fixed effects Poisson regression to perform the staggered difference-in-differences (DiD) analysis. The estimation results indicate that, on average, the launch of a mobile app leads to a 16.6% increase in invention patent applications and a 16.5% increase in patent citations. The dynamic DiD analysis reveals that the impact of mobile app launches on corporate innovation outcomes emerges only in the year following the

app's release and persists in subsequent years. The absence of treatment effect prior to the launch of the app strengthens our confidence that the parallel trend assumption is not violated.

To ensure the treatment effect is causal, we conduct four further analyses. First, we use high-dimensional interactive fixed effects to further alleviate the potential influences of contemporaneous events at the industry or region level (Chang et al., 2019; Kong et al., 2022) and reach the same conclusion. Second, the decision to launch an app is not exogenous, and the estimated treatment effect may be biased if the factors influencing app launches are correlated with those affecting innovation outcomes. To mitigate this selection bias, we follow He et al. (2024) to conduct a stacked-cohort DiD test based on a propensity score matching (PSM) sample. We continue to find a positive treatment effect with an economic magnitude similar to that of the baseline DiD regression. Third, to mitigate concerns that staggered DiD estimates may be biased due to treatment effect heterogeneity (Baker et al., 2022), we adopt alternative estimation methods proposed by Sun and Abraham (2021) and Borusyak, Jaravel, and Spiess (2024), consistently finding a positive treatment effect.

Besides innovation intensity and quality, we also explore the app effect from the perspective of efficiency. We expect that the large volume of personal data from app users can help companies identify innovation directions more quickly and accurately, enabling purposeful R&D investments. Additionally, using user data can help companies reduce costs associated with relying on external data for research. Consequently, this will improve the company's innovation efficiency. We use two proxies to measure innovation efficiency, i.e., patents per R&D expenditure (Hirshleifer, Hsu, and Li, 2013) and patent approval rate (Liu et al., 2024). We document that firms that launched mobile apps have enhanced innovation efficiency.

If the launching of apps promotes corporate innovation by collecting and utilizing user data, the positive effect should be stronger for firms with access to more such data. To test this

prediction, we first use two measures as proxies for the number of app users: (1) the number of cumulative downloads and (2) the number of people participating in app scoring (reviewers). We find that the treatment effect is more pronounced when apps have a larger user base. Second, following Wu (2023), we identify whether an app directly collects user-linked data from its data collection policy. Our findings reveal that the positive effect of app launching on corporate innovation is significant only for apps that gather user-linked data directly. Third, due to its nature, digital innovation relies particularly on intensive data (Beraja, Yang, and Yuchtman, 2023; Chang et al., 2024). If a company utilizes app user data to drive innovation activities, we expect it to achieve more digital innovation. To demonstrate that companies engage in more digital innovation after launching their apps, we calculate the share of a company's digital patents, measured as the proportion of digital patent applications to total patent applications in a given year. Firms generate a higher percentage of digital innovation outputs following the release of their apps.

The growing popularity of mobile apps raises public concerns about data privacy issues (Bian, Ma, and Tang, 2021). A relevant question in our context, then, is whether the innovation success driven by apps comes at the expense of consumer privacy or can be balanced with it (Deming, 2021). While a full analysis of this potential tradeoff and its welfare implications is beyond the scope of our paper, we attempt to offer two insights from the perspectives of app regulatory compliance and user-centric innovation. First, leveraging China's unique regulatory framework—where the Ministry of Industry and Information Technology (MIIT) publicly discloses firms' noncompliant data practices—we classify apps as either compliant or noncompliant. We find that both types of apps are positively associated with innovation outcomes and their marginal effects are statistically indistinguishable. Our finding provides preliminary evidence that firms can achieve innovation through access to customer data without necessarily breaching data privacy regulations. Second, we investigate whether

companies leverage data gathered through mobile apps to drive consumer benefits via innovation. Specifically, we examine the connection between app launches and user-centric innovations, such as those implemented in end-user scenarios or designed to enhance the user experience. By focusing on the composition of patent applications and citations, we provide consistent evidence that firms that launch mobile apps have a higher fraction of user-centric innovation compared to other types of innovation.

Our paper contributes to two strands of literature. First, while prior literature focuses primarily on the impact of incentives and finance (see He and Tian, 2018 for a review) and institutions (see He and Tian, 2020 for a review) on corporate innovation, we join the emerging literature on the role of information accessibility. Kong et al. (2022) show that information accessibility via the search engine plays a vital role in innovation. Hsu et al. (2024) find that pharmaceutical firms learn from the clinical study results mandatorily disclosed by peer firms in making their R&D decisions. Focusing on digital innovation, Beraja, Yang, and Yuchtman (2023) emphasize the value of access to government data for AI firms via government contracts with governments. While these studies focus on the value of public or government data, our paper differs from those studies by focusing on corporate internal data, i.e., the customer data accessed via mobile apps. In particular, app data differs from public data in its unique characteristics, enabling it to drive corporate innovation through distinct mechanisms, especially by enhancing commercial viability.

Second, our paper also contributes to the literature on mobile apps in different ways. Beyond the marketing literature (see Stocchi et al., 2022 for a review), accounting and finance scholars have recently explored the economic implications of apps for corporations. Wu (2023) develops a measure of mobile app value based on market reactions to app releases and demonstrates its ability to predict lower firm risk and higher future growth. Chen, Liu, and Wu (2024) show that mobile app downloads serve as a leading performance indicator, significantly

predicting future earnings, yet market participants underestimate their valuation impact. Ferracuti et al. (2024) find that following app adoption, management guidance becomes more accurate, and firms better manage both over- and under-investment. We extend this strand of literature by focusing on how apps stimulate corporate innovation in terms of intensity, quality, and efficiency.

Moreover, existing studies primarily emphasize the value of mobile data for firms, while largely overlooking its implications for consumers—the very providers of the data. Different from these studies, we show that firms that utilize customer data produce more consumer-oriented innovative outcomes. This suggests that, as data-driven innovation becomes more effective, the resulting innovations better meet consumer needs. Consequently, the benefits generated from data-driven innovation can be redistributed between firms and customers, thereby enhancing the value that consumers derive from their own data.

Finally, our study also complements Ferracuti et al. (2024) in exploring the value of app customer data for corporations in business operations and its relation to privacy regulation. While Ferracuti et al. (2024) show that privacy regulations limit the extent to which firms can leverage app data to enhance operational efficiency, our findings demonstrate that firms do not necessarily need to compromise regulatory compliance to drive innovation. Instead, firms can achieve innovation without breaching data privacy regulations.

2. Institutional Background, Features of Mobile Data, and Hypothesis Development

2.1. Background of Mobile Apps in China

Due to its massive population base and rapid economic development, China's mobile app market has shown substantial growth momentum and development potential in recent years. According to data from Newzoo, as of 2021, China had over 950 million smartphone users—more than the combined total of India, the United States, and Indonesia, which ranked second,

third, and fourth, respectively.[5] The rise in smartphone popularity has been accompanied by the popularity of mobile apps. In 2021, the number of active mobile devices in China surpassed 1.1 billion, with an average daily app usage time of 5.26 hours.[6] This reflects the degree of reliance on apps among Chinese users and the prevalence of mobile applications in daily life.

The statistics from MIIT show that as of 2021, the Chinese market had 2.515 million apps, including 1.348 million in Apple's App Store (China) and 1.167 million in local third-party stores.[7] Games were the largest category (709,000), followed by utility tools (370,000), e-commerce (248,000), lifestyle services (211,000), social communication (209,000), education (172,000), themes & wallpapers (124,000), news & reading (108,000), health & fitness (98,000), system tools (93,000), music & video (63,000), finance (55,000), photography (38,000), and other categories (17,000). The cumulative total downloads from third-party stores reached 2,107.2 billion, with game apps alone at 331.4 billion. Utility tools, music & video, and social communication apps had over 240 billion downloads each, while lifestyle services, news & reading, system tools, and e-commerce apps ranged from 140 to 196 billion. Finance and photography apps had 97.6 billion and 81.9 billion downloads, respectively. Worldwide, China has led the world in app downloads since 2021. Sensor Tower's State of Mobile 2024 report shows that in 2023, China led global mobile app downloads with over 113 billion, followed by India with 26.4 billion and the United States with 12.58 billion.[8]

Apps provide businesses with direct and timely interaction with users, serving as a crucial channel for gathering customer information and behavioral preferences. At the same time, app users are often required to share personal information in exchange for a better user experience. This information may include users' basic profiles, contact details, health and fitness, financial

[5] Source: https://tech.ifeng.com/c/8E2MgqH5zWm.
[6] Source: https://www.iresearch.com.cn/.
[7] Source: https://www.miit.gov.cn/hlwnj2021/hlw.html.
[8] Source: https://sensortower.com/state-of-mobile-2024.

information, transaction records, location data, device identifiers, and app usage logs (such as search history, browsing history, favorites, and social interactions).

Overall, mobile applications enable businesses to collect real-time, fine-grained customer data, which is difficult or impossible to obtain outside these platforms. Additionally, based on this study's statistics, for most A-share listed companies, their apps are not merely used to sell products but are integrated with their products and services, such as household appliances and automobiles. Therefore, user behavior data from these apps is crucial for companies to understand user feedback, gather customer preferences, improve product and service quality, and build competitive advantages. To provide a clearer picture of Chinese apps and the data they collect, Appendix 2 includes examples of typical apps from the sample companies examined in this study, covering corporate information and app information (including name, launching date, category, distribution area, download platforms, cumulative downloads, and the information collected).

2.2. Features of Customer Data on Apps

For firms, customer data collected through apps possesses unique characteristics compared to other types of data. In this study, we have summarized these features in three dimensions: relevance, proprietorship and exclusivity, and cost-effectiveness.

Relevance Customer data is closely and directly tied to the products and services offered by firms. As Veldkamp (2023) observes, much consumer data originates from transactions, which discloses not only customers' preferences and their willingness to pay but also a broad spectrum of behavioral and demographic traits. Because customer data is closely tied to a firm's products and services, they are not easily replaceable with data from other sources.

In particular, consumer data collected via proprietary platforms—such as mobile apps—is highly relevant to a firm's offerings, as it captures real-world interactions and experiences with its products or services. On apps, every user interaction—clicks, scrolls, inputs—is

automatically recorded. Each stage of the customer journey, from search to purchase and post-sale feedback, produces a rich stream of behavioral, contextual, and attitudinal data. For example, (1) search queries, keyword clicks, and time spent on pages can indicate emerging consumer needs; (2) product IDs, prices, discounts, payment methods, and purchase frequency provide insights into purchasing decisions; (3) textual reviews, ratings, customer service conversations, and return records reflect product usage experiences and satisfaction.

Traditionally, firms have relied heavily on organizational experience, managerial intuition, and market research. Under this paradigm, consumer data has played a relatively limited role in directly guiding decision-making and innovation activities, as they were often fragmented, retrospective, or lacking sufficient contextual richness. In contrast, the product- and service-specific data generated through apps provides real-time, granular insights into consumer behavior and preferences. This makes them not only valuable for understanding customer needs but also crucial inputs for innovation and predictive analytics.

Proprietorship and exclusivity Customer data collected via firm-owned digital platforms (e.g., mobile apps) is typically proprietary, meaning it is owned, managed, and controlled exclusively by the firm itself. Unlike syndicated data (e.g., general data provided by market research firms) or publicly available statistics, app-based customer data is deeply embedded in the firm's own business processes and customer interactions. This ensures high specificity and strong alignment with the firm's value proposition (Davenport and Redman, 2020).

Moreover, such data is also exclusive in that it is not accessible to external parties or competitors. Because it is generated through proprietary platforms, even firms offering similar products cannot replicate the specific customer interactions captured by the original firm. Such exclusivity provides firms with a sustainable competitive advantage, serving as both a strategic asset and a barrier to imitation (Du et al., 2021). The real-time and context-specific nature of customer data further reinforces its exclusivity. For instance, a smart home app may record

exactly when users turn on lights or adjust air conditioning—data derived from real-life usage scenarios that cannot be replicated through external sources. Since competitors cannot easily access or reconstruct such data, it becomes a unique and irreplaceable resource for driving innovation and strategic decision-making. Overall, these characteristics make customer data not only highly valuable but also extremely difficult to imitate.

Cost-effectiveness Firms typically do not build mobile apps solely to collect customer data. Most apps are developed to support core business functions—such as (1) facilitating sales and transactions (e.g., e-commerce apps), (2) assisting with product or service usage (e.g., companion apps for appliances or automobiles), or (3) functioning as standalone digital products themselves (e.g., gaming apps). As a result, customer data collected via firm-owned mobile apps is often generated as a natural byproduct of regular economic activities rather than the result of deliberate data-gathering initiatives.

Unlike traditional data collection methods—such as market research surveys or purchasing third-party data—this approach to data acquisition demands little additional effort or investment. Since apps are already in use for operational purposes, the cost of collecting data is effectively amortized or embedded in ongoing operations. Moreover, the production of customer data is fundamentally different from R&D-driven outputs like patents and technologies. Large datasets are not created in laboratories and do not require intensive labor (Veldkamp, 2023). Aside from a modest investment in skilled personnel to maintain digital infrastructure, the marginal cost of collecting customer data is nearly zero.

Additionally, customer data possesses characteristics such as large scale, real-time generation, accuracy, and continuous accumulation. These features further enhance its cost-effectiveness and reinforce its strategic value for business insight and innovation.

2.3. Hypothesis Development

Data serves as a critical input in the research and development of new products, and this represents one of the most important applications of AI in modern business (Babina et al., 2024). In particular, consumer data collected through mobile apps plays a pivotal role in influencing firm innovation through two key mechanisms: technological feasibility and commercial viability (Chatterjee et al., 2023).

Technological feasibility refers to whether a new technological solution can be effectively implemented and function as expected in real-world conditions. It concerns questions such as: Is the technology mature enough? Can algorithms or models operate reliably? Will the solution handle complex, dynamic environments? App-based customer data plays a critical role in addressing these questions. As summarized above, this data is typically large-scale, real-time, and contextualized. Because of these features, app-based customer data is ideal for training, testing, and optimizing digital technologies. It supports machine learning model development, facilitates A/B testing, and enables iterative improvements through continuous feedback loops. This helps firms avoid deploying technologies that are unreliable, poorly calibrated, or misaligned with actual user behavior.

In many digital innovations—such as recommendation algorithms, smart device functions, or user interface design—technical feasibility depends on access to massive amounts of high-quality data for training models (Cui and Curry, 2005; Bessen et al., 2022). Beraja et al. (2023) document that Chinese AI firms have significantly optimized their commercial product algorithms after obtaining facial data collected by public security departments. Real-world interaction data from apps reflects how customers actually behave, rather than relying on assumptions or simulations. As a result, companies can experiment more efficiently, detect system flaws earlier, and accelerate the improvement of technical solutions. For example, consider autonomous driving technology. To make such systems viable, companies need vast amounts of real-world data, including driving images, sensor readings, and environmental

conditions. These are collected through on-road usage via connected vehicles and mobile apps. Events like manual overrides, sudden braking, or route corrections are all automatically logged and analyzed. These real-time data streams are essential for training models to recognize and respond to complex scenarios such as rain, traffic congestion, or construction zones. In this case, the technological feasibility of autonomous driving hinges on whether the system has been trained and validated with enough high-quality, context-rich data. Without such data—generated from actual use—the system cannot reliably function. Therefore, app-based customer data transforms technological feasibility from a theoretical challenge into a testable, improvable process.

Confirming the technical feasibility is a necessary but insufficient condition for successful innovation. Equally important is establishing its commercial viability, i.e., market assessment, which refers to whether customers will actually adopt, use, and pay for the product or service (Kong et al., 2022). As highlighted by Keeley et al. (2013), successful innovation requires a "deep understanding of whether customers need or desire that invention." Commercial viability determines the market value of an innovation and directly impacts firms' return on investment and risk exposure. Without this market validation, even technically sound innovations may fail commercially.

Customer data collected via firm-owned apps plays a critical role in enhancing commercial viability, thereby encouraging firms to pursue innovation more confidently and effectively. These data—such as users' search behaviors, purchase histories, feature usage patterns, and feedback—provide rich, real-time, and relevant insights into customer preferences and evolving needs. This supports innovation through three key directions.

First, app-based customer data helps firms identify latent market needs and generate new innovation ideas. Unlike traditional market research, which is often retrospective and aggregated, app data captures real-time user intent and micro-level engagement. For instance,

when users frequently search for certain features that do not yet exist or repeatedly use a workaround to achieve specific outcomes, these behavioral signals may indicate unfulfilled needs. Without such data, firms might misinterpret low sales as a lack of interest, missing opportunities for innovation. Customer data enables firms to detect these subtle cues early, thereby refining R&D priorities and aligning innovation with real demand.

Second, customer data significantly reduces market uncertainty and innovation risk (Wu, 2023; Veldkamp, 2023). Market interest uncertainty is an important factor restricting corporate innovation. As digitized information, data serves as a predictive tool that helps firms make informed judgments about whether a product or feature will succeed in the market. Through techniques such as A/B testing, behavioral analysis, and demand forecasting, firms can pre-assess customer reactions and optimize features before committing substantial resources. This improves resource allocation, enhances decision quality, and reduces the likelihood of costly innovation failures.

Third, app-based customer data enhances commercial viability by enabling product and service customization (Kopalle, Kumar, and Subramaniam, 2020). A one-size-fits-all approach often fails in today's fragmented and competitive markets. By analyzing detailed data such as usage frequency, preferred features, and demographic characteristics, firms can identify user segments and develop personalized offerings—such as tailored recommendations, dynamic pricing, or targeted promotions. This customization boosts product relevance and user satisfaction, leading to higher engagement and retention. From an innovation perspective, this ensures that new offerings are precisely aligned with diverse customer preferences, increasing their likelihood of market success.

In summary, app-based customer data allows firms to evaluate whether innovations are not only technically viable but also aligned with market demand. This dual validation strengthens innovation outcomes by ensuring that new products or features are both feasible to

develop and appealing to users. Building on the above analysis, we propose the following research hypothesis.

H1a: The release of a mobile app fosters corporate innovation.

Nevertheless, we also note that extensive access to customer data via apps may not necessarily lead to enhanced innovation performance and can even have adverse effects. The recent study by Hoelzemann et al. (2024) introduces a theoretical framework termed the "streetlight effect," illustrating how the easy availability of data on moderately promising projects can inadvertently constrain exploration and dampen innovation. Their model suggests that when data illuminates projects that appear attractive but are ultimately suboptimal, it can lead agents to focus disproportionately on these options, thereby narrowing the scope of exploration and reducing the likelihood of breakthrough discoveries. In our context, the "streetlight" effect can crowd out more innovative pursuits that lack immediate data support. For instance, if app analytics highlight certain user behaviors, companies might prioritize features catering to these behaviors, neglecting other areas that could offer greater innovation potential but lack immediate data representation. Relatedly, by analyzing customer behavior data, companies can more effectively target their desired markets, optimize product pricing, and enhance customer experiences, leading to rapid short-term growth. However, this reliance on existing data strategies may cause companies to focus excessively on current market demands, thereby neglecting or disincentivizing the exploration of future trends and innovation. In addition, relying on app users' data for innovation may suffer from sampling bias. App-derived data often reflects the behaviors and preferences of the most active users, which may not represent the broader customer base. As highlighted by Cao et al. (2024), an abundance of user data does not inherently lead to higher-quality innovation when the data is biased and unrepresentative. Given the arguments outlined above, we form the alternative hypothesis as:

H1b: The release of a mobile app has no impact or negatively affects corporate

innovation.

3. Empirical Design

3.1. Data and Sample

We use a sample that combines financial data of public firms from CSMAR, patent data from CNRDS, and manually collected mobile app data. We obtain the names, operators, release dates, and other information of mobile apps from Qimai Data (https://www.qimai.cn/).[9] We manually match the names of listed companies and their subsidiaries with the names of app operators.

We start with all Chinese A-share listed firms from 2007 to 2021. The sample begins in 2007 because all listed firms were required to disclose the R&D information after the implementation of new accounting standards in that year. As there is, on average, a two-year lag between the date when inventors file for patents (the application date) and the date when patents are granted, we end our sample period in 2021 to account for this issue.

We apply the following criteria to construct our sample for empirical analysis[10]. (1) We exclude firms in the financial industry. In addition, companies in the software and related services industry may develop apps not for their own use but for other companies. Therefore, we also exclude companies from this industry. (2) Because some industries rarely use mobile

[9] Qimai Data currently covers App Store and Google Play data from 155 countries/regions worldwide, as well as data from nine major Android app markets in China. As of December 31, 2021, it has recorded data on over 13 million apps and more than 20 million keywords, with a total query volume exceeding 400 million.
[10] See Appendix 3 for a detailed description of the sample changes at each step.

apps for business purposes, such as agriculture, we follow Ferracuti et al. (2024) and retain only firms in industries where at least 5% of firms have launched mobile apps during the sample period. (3) We restrict firms in the treatment group[11] to have observations before and after the launching of the app to ensure that we can compare the changes in innovation output. In this step, firms that release their mobile apps before listing are excluded. (4) Following the practice in the literature (e.g., Kong et al. 2022; Chang et al. 2015), we focus on a group of innovative firms with at least one eventually granted invention patent filed and cited during the sample period. This requirement helps ensure that the sample firms are actively engaged in innovation. (5) We exclude observations with missing values on the variables and firms with special treatment (ST or *ST) determined by stock exchanges. Our final sample encompasses 24,661 firm-year observations from 2,582 listed firms, including 308 firms with mobile apps.

As shown in Panel A of Table 1, 760 firms in the initial sample have launched their mobile apps, and the first app in the sample was released in 2009. We also analyse the industry distribution of the final sample and find that firms in most industries have released their mobile Apps. As shown in Panel B of Table 1, the proportions are relatively higher in the manufacturing, wholesale and retail trade, and real estate industries. This also indicates that our research findings are not driven by a few specific industries.

[Insert Table 1 About Here]

3.2. Variables

Innovation Output The number of corporate patent applications is widely used as a measurement of firm innovation output. Patent application data can represent corporate

[11] We define the "treatment group" as firms that launched their first mobile app at any time during our sample period.

innovation activities more objectively and comprehensively compared to R&D investment intensity (He and Tian, 2013). According to patent law, patents can be classified into three types: invention patents, utility model patents, and design patents. Among these, invention patents are granted for innovative technical solutions involving products, processes, or improvements and undergo more stringent substantive examination, reflecting higher standards of inventiveness and longer protection terms (Han, Jiang, and Mei, 2024). Therefore, we focus on invention patents in this study. We include patent applications filed by the sample firms and their subsidiaries to measure patenting outcomes.

Following the innovation literature, we construct two measures to gauge a firm's innovation output. The first measure is the total number of invention patent applications filed by a firm in a given year that are eventually granted (*Pat*). We use the patent application year rather than the grant year, as the application date is closer to the actual time of invention (Hall, Jaffe, and Trajtenberg, 2001). The second measure is constructed by counting the number of non-self-citations to all eventually granted invention patents filed by a firm in a given year (*Cite*). A patent's citations reflect the quality and the impact of the invention (Hall, Jaffe, and Trajtenberg, 2005). We exclude citations made by a firm itself, as self-citations may not accurately capture the broader impact of the firm's innovation on others (He and Tian, 2013; Kong et al., 2022).

The patent data are subject to a truncation problem. Due to the application-grant lag, patent applications filed toward the end of our sample period are less likely to be granted within our sample period, leading to underrepresentation in the data.[12] To address this issue, we adjust the raw patent count data by using the "weight factors" computed from the application-grant lag distribution. This adjustment follows the guidance of Hall, Jaffe, and Trajtenberg (2001, 2005)

[12] For instance, a patent applied for in 2021 might not be granted until 2024. However, since the patent data we collected is only updated until 2023, patents granted in 2024 are not covered.

and is a common practice in innovation studies (Tian and Wang, 2014; Kong et al., 2022).[13] Similarly, due to sample-end censoring, patents in the latter years of the sample have less time to accumulate citations. We correct this bias by using the fixed effects approach suggested by Hall, Jaffe, and Trajtenberg (2001) and widely used in prior studies (Chang et al., 2015; Kong et al., 2022; Chang et al., 2019). Specifically, we divide the raw value of the number of citations by the average number of citations in the same technological class and the same application year. We then aggregate the adjusted number of citations for all patents filed by each firm each year to obtain the final number of citations.[14] All innovation variables are winsorized at the 1st and 99th percentiles by year to mitigate the influence of outliers.

Applaunch We measure firms' mobile app adoption using *Applaunch*, an indicator variable set to one in the year of and years following the initial mobile app launch, and zero otherwise. Following Ferracuti et al. (2024), we use the year of the first app release for firms with multiple apps. We define the "treatment group" as firms that launched their first mobile apps at any time during our sample period. Firms that never launched a mobile app by the end of the sample period (i.e., 2021) are defined as the "control group".

Control Variables We control for a range of firm characteristics that have been documented as important determinants of innovation by previous studies (Kong et al., 2022; Chang et al., 2015; Tian and Wang, 2014). These control variables include firm size (*Size*), the

[13] Specifically, we adjust the number of patents as follows:

$$Pat_{i,t}^{adj} = \frac{Pat_{i,t}^{raw}}{\sum_{k=0}^{endyr-t} frac_k}$$

$Pat_{i,t}^{raw}$ and $Pat_{i,t}^{adj}$ are the numbers of patents of firm *i* in year *t* before and after adjustment, respectively, with *t* being the application year. $frac_k$ denotes the fraction of patents granted during the kth year since their application, estimated for the 1990-2014 period. *endyr* is the last year of the sample (i.e., 2021). The denominator then represents the fraction of successful patent applications in year t that are expected to be granted by the end of our sample, based on historical trends. Therefore, we can infer the "latent" total number of patents that are or will be eventually granted even beyond our sampling period.

[14] We use several alternative citation-based innovation measures. Firstly, we calculate the average number of forward citations to each eventually granted patent filed by a firm in a given year to measure the impact of innovation at the intensive margin. Second, we adjust the raw citation counts by dividing the raw number of citations by the fraction of predicted lifetime citations based on a citation-lag distribution (Tian and Wang, 2014; Brav et al., 2018; Xu et al., 2024). We get the similar results when using alternative measures.

book-to-market equity ratio (*BM*), leverage (*Leverage*), cash holdings (*Cash*), profitability (*ROA*), firm age (*Age*), institutional shareholding (*InstOwn*), ownership concentration (*OwnCon*), tangible assets (*PPE*), and research and development expenditure (*R&D*). In addition, we include the Herfindahl-Hirschman Index (*HHI*) and its squared item (HHI^2) at the industry level as control variables to account for the impact of competition (Aghion et al., 2005). Detailed definitions of all variables are provided in Appendix 4. To mitigate the influence of outliers, all control variables are winsorized at the 1st and 99th percentiles.

Table 2 provides the descriptive statistics of the key variables. Based on measures of innovation, our sample firms on average file 23 invention patents (eventually granted) and receive 11 citations from other firms each year. The median numbers of patents and citations are 3.4 and 1.6, consistent with the highly right-skewed pattern of innovation outcomes. In addition, an average firm has a size measure of 8.346 in terms of log value and RMB 12,600 million in total assets, a book-to-market equity ratio of 0.445, and an age of 10.5 years. The average firm holds a sizeable amount of cash with a cash ratio of 19% of the total assets and is moderately levered with a book leverage ratio of 43.1%. Regarding performance, a sample firm has an average ROA ratio of 3.9%. Regarding ownership, the average firm has an institutional shareholding ratio of 44.9%, and the shareholding of the largest shareholder is 34.7%, indicating a high level of ownership concentration. On average, R&D accounts for 1.8% of total assets. Overall, the summary statistics of the key variables in our sample are comparable to prior studies (e.g., Kong et al., 2022; Dai et al., 2025).

[Insert Table 2 About Here]

4. The Effect of Mobile Apps on Corporate Innovation

4.1. Baseline Results

To test our hypothesis, we employ a DiD model for the baseline analysis, using the following specification:

$$Innovation_{i,t+1} = \alpha + \beta Applaunch_{i,t} + \gamma \boldsymbol{Z}_{i,t} + \varphi_i + \delta_t + \varepsilon_{i,t} \quad (1)$$

where *i* indexes firm and *t* indexes year. The dependent variables, *Pat* and *Cite*, capture corporate innovation. $\boldsymbol{Z}_{i,t}$ is a vector of firm and industry characteristics that could affect a firm's innovation activity, as previously discussed. φ_i represents firm fixed effects, and δ_t represents year fixed effects. Standard errors are clustered at the firm level. When estimating models with count variables as dependent variables, such as patenting activities, the log-linear regressions may produce estimates that lack meaningful interpretation. In contrast, Poisson regression provides estimates with more natural interpretations (Cohn, Liu, and Wardlaw, 2022; Chen and Roth, 2024). Consequently, following recent studies (He et al., 2024; Heller, 2024), we employ fixed effects Poisson regressions for our main analyses.

We report the baseline results in Table 3. The coefficient of *Applaunch* in column (1) is significantly positive at the 5% level, indicating that companies file more patent applications after launching their apps. Furthermore, in column (2), the coefficient of *Applaunch* is positive and statistically significant at the 1% level, suggesting that the quality of innovation also improves notably after the app release, as evidenced by the significant increase in patent citations. We also interpret the economic implications of the coefficient estimates. We find that, on average, the initial release of a mobile app is associated with a 16.6% increase in patent applications and a 16.5% increase in patent citations.[15] This shows that the impact of app launches on corporate innovation activities is not only statistically significant but also economically meaningful. Overall, the baseline regression results support the prediction of Hypothesis 1a.

[15] The Poisson model takes the form $E[y|\boldsymbol{x}] = \exp(\boldsymbol{x\beta})$. For coefficient interpretation, a one-unit increase in X_i, y is multiplied by e^{β_i}. In our patent setting, as column (1), the value of *Applaunch* goes from 0 to 1 multiplies the expected number of patents by $e^{0.154}$=1.166, meaning a 16.6% increase.

[Insert Table 3 About Here]

4.2. Robustness Tests

4.2.1. Dynamic Analysis

The key assumption of DiD estimations requires that in the absence of the event, the average change in patents would have the same trends for the treatment and control groups. To test the validity of this assumption in the pre-treatment period, we examine the dynamic effect of innovation activities around the app launch by estimating the following equation:

$$Innovation_{i,t+1} = \alpha + \beta_1 Applaunch(-3^{+})_{i,t} + \beta_2 Applaunch(-2)_{i,t} + \beta_3 Applaunch(0)_{i,t} + \beta_4 Applaunch\ (+1)_{i,t} + \beta_5 Applaunch\ (+2^{+})_{i,t} + \gamma \boldsymbol{Z}_{i,t} + \varphi_i + \delta_t + \varepsilon_{i,t} \quad (2)$$

We defined five dummy variables, *Applaunch (-3+)*, *Applaunch (-2)*, *Applaunch (0)*, *Applaunch (+1*), and *Applaunch (+2+)*, to indicate the years relative to the event year. *Applaunch (-3+)* is a dummy variable that takes the value of one if the observation occurs three or more years before the release of the mobile app, and zero otherwise. *Applaunch (-2)* is a dummy variable equal to one if the observation occurs in the second year prior to the app's release, and zero otherwise. *Applaunch (0)* is set to one if the observation falls within the year of the app's release, and zero otherwise. *Applaunch (+1*) is a dummy variable that equals one for the first year following the app's release, and zero otherwise. Finally, *Applaunch (+2+)* is a dummy variable that takes the value of one for observations occurring two or more years after the app's release, and zero otherwise. The year before the release of the mobile app is the reference year, and its corresponding dummy variable is omitted from the regression to avoid multicollinearity. All other variables are the same as in the baseline DiD regression.

We report the results of the dynamic analysis in Table 4. The coefficients β_1 and β_2 are not statistically significant across both columns. This indicates no significant pre-trends in innovation activities between the treatment and control groups, providing greater confidence that the parallel trend assumption of the DiD is not violated. This also implies that the positive

effect of app launching on innovation is unlikely to be confounded by unobservable differences between the treatment and control firms. Furthermore, by examining the coefficients associated with post-event time dummies, we find that the β_4 and β_5 are consistently significant at conventional levels. This implies that the positive effect of app release generally begins in the second year and persists beyond (the dependent variables are measured in t+1). This finding is consistent with the fact that innovation efforts require time to result in patents.

[Insert Table 4 About Here]

4.2.2. Controlling for Interactive Fixed Effects

We further replace the fixed effects with interactive fixed effects to mitigate the potential influences of contemporaneous events at the industry or region level (Chang et al. 2019; Kong et al. 2022). Specifically, we include industry-year fixed effects to control for potential differential trends in patenting activities and app release across industries over time. We also include province-year fixed effects to capture unobserved, time-varying region-level factors. Table 5 indicates that our main results remain robust after incorporating industry-by-year fixed effects, province-by-year fixed effects, or both.

[Insert Table 5 About Here]

4.2.3. Stacked-cohort DiD Tests Based on A Matched Sample

The decision for a company to launch an app could be influenced by many factors. To the extent that these factors are also correlated with corporate innovation, our estimated effect of app launching on innovation is subject to selection bias. To mitigate endogeneity concerns stemming from non-random treatment assignments, we follow He et al. (2024) to conduct stacked-cohort DiD tests based on the propensity score matching (PSM) sample.

Specifically, instead of estimating propensity scores using a pooled sample over time, we run separate propensity score models on a year-by-year basis to ensure that matching is performed by observation year (Boyd, Kannan, and Slotegraaf, 2019). For each event year *t*, we construct a control group comprising firms that have never adopted mobile apps during the period from year *t*-4 to year *t*+4. We employ a one-to-three nearest-neighbor propensity score matching method to select three control firms from the same industry for each treatment firm. The propensity score is estimated using a logit model with predictors that include firm characteristics, as well as the natural logarithm of one plus the number of patents that a firm has applied for and was later granted (*LnPat*). All predictors are measured as the average value in year *t*-1 to year *t*-3. Each treatment firm is matched to its control firms to form an event cohort. In each cohort, we include firm-year observations from four years before to four years after the event year. We then stack all cohorts to form our regression sample, which comprises 8,828 cohort-firm-year observations.

We first examine the matching results and present the results in Panel A of Table 6. The p-statistics for univariate comparisons between treatment and control firms' pre-event characteristics after matching are all greater than 0.1, indicating that the propensity score matching process has successfully formed a control group that is highly similar to the treatment group.

Next, we estimate the average treatment effect of app release on corporate innovation across all cohorts using the following equation:

$$Innovation_{c,i,t+1} = \alpha + \beta Applaunch_{c,i,t} + \gamma \boldsymbol{Z}_{c,i,t} + \varphi_{c,i} + \delta_{c,t} + \varepsilon_{c,i,t} \quad (3)$$

We include cohort-firm fixed effects ($\varphi_{c,i}$) and cohort-year fixed effects ($\delta_{c,t}$) to control for firm and time heterogeneity within each cohort, respectively. Thus, our stacked-cohort DiD specification exploits within-cohort-firm variation in innovation activities resulting from app launches. All other variables are the same as in Equation (1).

The results of stacked-cohort DiD tests are presented in Panel B of Table 6. The coefficients of *Applaunch* remain significantly positive, and their economic magnitude is similar to that of the baseline regression. Overall, these results provide confidence that our findings are unlikely to be influenced by differences in characteristics between the treatment and control groups.

[Insert Table 6 About Here]

4.2.4. Heterogeneity-robust DiD Estimation

Recent literature has highlighted that the traditional staggered DiD regression could produce biased estimates for the average treatment effect (ATE) due to treatment effect heterogeneity. To address this concern, we adopt alternative methods introduced by Sun and Abraham (2021) [16] and Borusyak, Jaravel, and Spiess (2024) to re-estimate the ATE, respectively. Since the estimators proposed by these studies are based on linear regressions, we use $LnPat_{t+1}$ and $LnCite_{t+1}$ as dependent variables and employ the OLS model for the estimations.

Table 7 reports the estimation results. Regardless of the DiD method used, the coefficients of *Applaunch* are significantly positive at the 1% or 5% level.[17] Overall, these results indicate that our main inference remains largely unchanged under alternative DiD methods.

[Insert Table 7 About Here]

4.3. Innovation Efficiency

Besides of focusing on innovation intensity and quality, we further examine whether acquiring user data helps companies engage in more effective innovation. We use two

[16] Sun and Abraham (2020) propose the interaction-weighted (IW) estimator as an alternative to the two-way fixed effects (TWFE) regression in the presence of treatment effect heterogeneity across cohorts.
[17] There are no single-regression R^2 to report when using the methods of Borusyak et al. (2024) because they aggregate separate group-time effect estimates.

approaches to measure innovation efficiency. The first measure captures a firm's ability to generate patents per R&D investment. We expect that the large volume of personal data from app users can help companies identify innovation directions more quickly and accurately, enabling purposeful R&D investments. Additionally, using user data can help companies reduce costs associated with relying on external data for research. Consequently, this will improve the company's innovation efficiency. Similar to Hirshleifer, Hsu, and Li (2013), we define patents per R&D expense ($LnPatR\&D_t$) as the natural logarithm of one plus the number of successful patent in year *t* divided by R&D expenses in year *t*-1 in million, $ln\left(1+\frac{patents_t}{R\&D_{t-1}}\right)$. Next, we consider the success rate of patent applications. If user data provides companies with more accurate information to assist in innovation decisions, the patents produced by the company will have higher feasibility and applicability. We expect that the success rate of patent application will significantly increase after the company launches the app. Similar to Liu et al. (2024), we measure the success rate of patent applications (*Approval*) as the ratio of granted patents over total patent applications. We test the relation between innovation efficiency and app release using the OLS regression model.

Table 8 reports the results. The positive and significant coefficients of *Applaunch* in both columns indicate the improvement of innovation efficiency following the release of a mobile app, evidenced by a higher likelihood of innovation success both in the R&D process and the application process.

[Insert Table 8 About Here]

4.4. Underlying Mechanism

In this study, we propose that the primary function of app customer data is to serve as a key data element in the innovation process. To evaluate this, we first analyze the varying impact

of app data scale on innovation. Next, we compare the effects of app launches on digital innovation versus other forms of innovation.

4.4.1. Heterogeneous Tests Based on Data Scale

If launching apps promotes corporate innovation by collecting and utilizing user data, the positive effect should be stronger for firms with access to more customer data. We test the proposed channel using the following regression specification:

$$Innovation_{i,t+1} = \alpha + \beta_1 Applaunch_{More_{i,t}} + \beta_2 Applaunch_{Less_{i,t}} + \gamma \boldsymbol{Z}_{i,t} + \varphi_i + \delta_t + \varepsilon_{i,t} \quad (4)$$

where *Applaunch_More* and *Applaunch_Less* are indicators for the scale of user data.

We identify the scale of user data available to the firm in two methods: (1) the number of app users, and (2) whether an app directly collects user-linked data. A larger app user base provides the company with a broader and more diverse range of data sources. As the number of users increases, so do the variety of usage scenarios and user behaviors captured, encompassing a wider demographic. This expansion enhances the diversity of data types—such as location, preferences, and consumption habits—leading to richer and more representative samples. With more users, the company can generate more accurate data analyses (e.g., by training more effective machine learning models) and gain deeper insights into user trends and market segments. Ultimately, this supports more informed and innovative decision-making.

We use two measures as proxies for the number of app users: (1) the number of cumulative downloads and (2) the number of people participating in app scoring (reviewers). We divide all apps in our sample into two groups based on the median of app download volume (or app reviewer volume).[18] If the app download volume (or app reviewer volume) is above the median

[18] The data we collect are the cumulative download volume and the cumulative review volume up to the data-collection date (i.e., January 2025). However, the raw volumes for apps launched in different years are not directly comparable due to varying time spans. To address this issue, we sort the sample based on the median value by the initial release year.

and the observation takes place after the release of the mobile app, then *Applaunch_More* is assigned a value of one; otherwise, it is assigned a value of zero. Conversely, if the app download volume (or app reviewer volume) is below the median and the observation takes place after the release of the mobile app, then *Applaunch_Less* equals one and zero otherwise.

To identify whether the apps in our sample directly collect user-linked data, we rely on several sources. First, following Wu (2023), we hand-collect information on each app's data collection policy from its homepage in the App Store. The app policy specifies the types of user-linked data collected, such as location, contact information, search history, and other related data. Second, we review the app's privacy agreement to identify whether user-linked data is directly collected. We divide all apps in our sample into two groups based on whether they collect user-linked data. If the app collects user-linked data and the observation takes place after the release of the mobile app, then *Applaunch_More* is assigned a value of one; otherwise, it is assigned a value of zero. Conversely, if the app does not collect user-linked data and the observation takes place after the release of the mobile app, then *Applaunch_Less* equals one and zero otherwise. In our sample, 212 apps collect user data directly, while the remaining 96 do not.

We present the results in Panel A of Table 9. Columns (1) and (2) display the results based on app download volume. No matter whether innovation activity is measured by the number of patents or the citation counts, the coefficient of *Applaunch_More* is significantly positive, whereas the coefficient of *Applaunch_Less* is insignificant. Columns (3) and (4) show similar results when using app reviewer volume as the basis. Columns (5) and (6) present the results based on whether apps directly collect user-linked data. We also find that the coefficient of *Applaunch_More* is significantly positive and larger in magnitude, while the coefficient of *Applaunch_Less* remains insignificant across different measures of innovation. These findings

suggest that companies are more likely to engage in innovation activities when they acquire more user data through their apps.

[Insert Table 9 About Here]

4.4.2. Data-driven Innovation (Digital Innovation)

Compared to traditional innovation, digital innovation is distinguished by its unique characteristics, including reprogrammability, data homogenization, and the self-referential nature of digital technology (Yoo et al., 2012). In line with these unique characteristics, data is the key driver of success in digital innovation (Beraja, Yang, and Yuchtman, 2023). Therefore, when a company leverages app user data to drive innovation, it is expected to generate more digital innovation. To demonstrate that companies engage in more digital innovation after launching their apps, we calculate the share of a company's digital patents, measured as the proportion of digital patent applications to total patent applications in a given year (*DigitalPat_shr*).[19] Similarly, we define the share of citations of digital patents relative to total patent citations as *DigitalCite_shr*. A patent is classified as a digital patent based on the presence of specific keywords in its abstract. The keyword dictionary used for this classification is provided in Appendix 5. We then regress the share of a company's digital patents on app release using an OLS model.

The test results are presented in Panel B of Table 9. Regardless of whether the share of digital innovation is measured by the proportion of digital patent applications (*DigitalPat_shr*) or citations (*DigitalCite_shr*), the coefficient of *Applaunch* is significantly positive at the 1% level. This finding suggests that companies generate more digital innovation outputs following the release of their apps.

[19] Considering that some observations may have a total patent count of zero in a given year, we add one to the total number of patents as the denominator.

4.5. Discussion on Data Value and User Privacy

The increasing prevalence of mobile apps has heightened public concerns regarding data privacy (Bian, Ma, and Tang, 2021). As companies increasingly rely on app-generated user data to drive innovation, a critical question arises: Does this innovation success come at the cost of consumer privacy, or can it be achieved while maintaining a balance between technological advancement and user protection (Deming, 2021)? This debate is particularly relevant in the digital economy, where user data serves as both a catalyst for innovation and a potential point of vulnerability. On the one hand, access to vast amounts of user data potentially enables companies to enhance product development, optimize user experiences, and create data-driven solutions that push the technological boundaries. On the other hand, concerns persist regarding how companies collect, store, and utilize this data—especially in the absence of robust privacy protections (Acquisti, Brandimarte, and Loewenstein, 2015). In this section, we add to this discussion by distinguishing apps based on their regulatory compliance and studying user-centric innovation.

4.5.1. The Regulatory Compliance of Apps

Leveraging user data via apps raises ethical and regulatory concerns regarding data security, informed consent, and potential misuse. A key question that arises is whether innovation success is driven at the expense of complying with data privacy regulations. We attempt to explore this issue by identifying the regulatory compliance of an app based on whether it has been criticized by regulatory authorities for illegally collecting or using users' data. We collect unique data on violations of privacy regulations from the website of the MIIT, with data coverage up to 2024.[20] Among 308 apps in our sample, 31 of them have been criticized by MIIT. We define *Applaunch_Comp* as equal to one if the app has not been

[20] The Ministry of Industry and Information Technology of China (MIIT) and its local branches regularly issue public notices and administrative penalties regarding the apps operated by enterprises on their official websites. For example, https://www.miit.gov.cn/xwdt/gxdt/sjdt/art/2024/art_424e1faa0479457689f7a49e113a5429.html.

criticized by MIIT and zero otherwise. *Applaunch_NonComp* equals one if the app has been criticized by MIIT and zero otherwise. All other variables are the same as in Equation (4).

The estimation results are presented in Table 10. We find that the coefficients of both *Applaunch_Comp* and *Applaunch_NonComp* are significant and positive. Moreover, although the coefficient of *Applaunch_Comp* is larger than that of *Applaunch_NonComp*, their statistical difference is insignificant. Therefore, there is no evidence to suggest that firms achieve superior innovation performance by violating data privacy regulations when collecting user data from their apps. Instead, companies can still drive meaningful corporate innovation while adhering to legal and ethical data practices.

[Insert Table 10 About Here]

4.5.2. Effect of User-centric Innovation

Another related question in the data privacy debate is how customers can benefit from sharing their data with firms via apps. In our context of corporate innovation, we aim to explore this question from the perspective of user-centric innovation, such as innovations applied in end-user scenarios or aimed at improving user experience. The increase in user-oriented innovation forms the foundation for companies to offer higher-quality products or services, ultimately benefiting consumers and improving consumer welfare. In today's digital landscape, mobile apps have emerged as a critical touchpoint between firms and users, offering new channels for interaction, feedback, and personalization. By enabling more direct and dynamic engagement with users, the introduction of mobile apps may foster user-oriented innovation, helping firms better understand and respond to evolving customer needs. Understanding this relationship is crucial for evaluating the broader implications of releasing mobile apps on innovation and consumer welfare.

To test whether companies engage in more user-centric innovation after launching their apps, we calculate the share of a company's user-centric patents, measured as the proportion of user-centric patent applications to total patent applications in a given year (*UserPat_shr*).[21] Similarly, we define the share of citations of user-centric patents relative to total patent citations as *UserCite_shr*. A patent is classified as user-centric based on the presence of specific keywords in its abstract. The keyword dictionary used for this classification is provided in Appendix 5. We then regress the share of a company's user-centric patents on app release using an OLS model.

The test results are presented in Table 11. When using the proportion of user-centric patent applications (*UserPat_shr*) as the dependent variable, the coefficient of *Applaunch* is significantly positive at the 1% level. When regressing the proportion of user-centric patent citations (*UserCite_shr*) as the dependent variable, the coefficient of *Applaunch* is positive and significant at the 10% level. These results suggest that companies generate more user-centric innovation outputs following the launch of their apps. This also indicates that the use of app-generated user data plays a key role in driving innovations that are closely aligned with consumer needs and preferences, consistent with our hypothesis that customer data provides directions for corporation innovation. Overall, we provide evidence that customers may benefit from sharing their data through apps, as it enables companies to develop more personalized, tailored products and services that enhance the overall user experience.

[Insert Table 11 About Here]

5. Conclusions

[21] Considering that some observations may have a total patent count of zero in a given year, we add one to the total number of patents as the denominator.

In this study, we explore the role of app-based customer data in driving corporate innovation. This data is uniquely characterized by its relevance to a firm's products and services, its proprietary and exclusive nature, and its cost-effectiveness. Based on these features, we hypothesize that customer data from mobile apps stimulates corporate innovation by supporting both technological feasibility and commercial viability. Leveraging mobile app launches as quasi-natural experiments, we find strong evidence that access to app-generated data significantly boosts the intensity, quality, and efficiency of firm innovation. This effect is particularly pronounced for apps with a larger user base, those that collect user-linked data directly, and digital innovations that are highly data-dependent. Moreover, our findings suggest that regulatory compliance regarding consumer privacy does not significantly alter the innovation impact of app launches. Furthermore, firms tend to develop more user-centric innovations aimed at enhancing consumer experience. These insights underscore the crucial data element role that app-generated data plays in the corporate innovation process and contribute to the recent debate on the trade-off between technological advancement and data privacy concerns.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *The Quarterly Journal of Economics*, *120*(2), 701-728.

Aghion, P., & Jaravel, X. (2015). Knowledge spillovers, innovation and growth. *The Economic Journal*, *125*(583), 533-573.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*, 103745.

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370-395.

Beraja, M., Yang, D. Y., & Yuchtman, N. (2023). Data-intensive innovation and the state: Evidence from AI firms in China. *The Review of Economic Studies*, *90*(4), 1701-1723.

Bessen, J., Impink, S. M., Reichensperger, L., & Seamans, R. (2022). The role of data for AI startup growth. *Research Policy*, *51*(5), 104513.

Bian, B., Ma, X., & Tang, H. (2021). The supply and demand for data privacy: Evidence from mobile apps. *Working paper*.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, *91*(6), 3253-3285.

Boyd, D. E., Kannan, P. K., & Slotegraaf, R. J. (2019). Branded apps and their impact on firm value: A design perspective. *Journal of Marketing Research*, *56*(1), 76-88.

Brav, A., Jiang, W., Ma, S., & Tian, X. (2018). How does hedge fund activism reshape corporate innovation? *Journal of Financial Economics*, *130*(2), 237-264.

Cao, R., Koning, R., & Nanda, R. (2024). Sampling bias in entrepreneurial experiments. *Management Science*, *70*(10), 7283-7307.

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, *134*(3), 1405-1454.

Chang, X., Chen, Y., Wang, S. Q., Zhang, K., & Zhang, W. (2019). Credit default swaps and corporate innovation. *Journal of Financial Economics*, *134*(2), 474-500.

Chang, X., Fu, K., Low, A., & Zhang, W. (2015). Non-executive employee stock options and corporate innovation. *Journal of Financial Economics*, *115*(1), 168-188.

Chatterji, A. K., & Fabrizio, K. (2012). How do product users influence corporate invention? *Organization Science*, *23*(4), 971-987.

Chatterjee, S., Chaudhuri, R., Mariani, M., & Wamba, S. F. (2023). The consequences of innovation failure: An innovation capabilities and dynamic capabilities perspective. *Technovation*, *128*, 102858.

Chen, J., & Roth, J. (2024). Logs with zeros? Some problems and solutions. *The Quarterly Journal of Economics*, *139*(2), 891-936.

Chen, S., Liu, Y., & Wu, X. (2024). Performance indicators of the digital age: Mobile apps, firm disclosure, and stock returns. *Working paper*.

Cohn, J. B., Liu, Z., & Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, *146*(2), 529-551.

Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, *24*(4), 595-615.

Dahl, D. W., Fuchs, C., & Schreier, M. (2015). Why and when consumers prefer products of user-driven firms: A social identification account. *Management Science*, *61*(8), 1978-1988.

Dai, Y., Du, T., Gao, H., Gu, Y., & Wang, Y. (2024). Patent pledgeability, trade secrecy, and corporate patenting. *Journal of Corporate Finance*, *85*, 102563.

Davenport, T. H., & Redman, T. C. (2020). Your organization needs a proprietary data strategy. *Harvard Business Review*, 2-6.

Deng, S., Mao, C. X., & Xia, C. (2021). Bank geographic diversification and corporate innovation: Evidence from the lending channel. *Journal of Financial and Quantitative Analysis*, *56*(3), 1065-1096.
Deming, D. (2021). Balancing privacy with data sharing for the public good. *The New York Times*.
Du, R. Y., Netzer, O., Schweidel, D. A., & Mitra, D. (2021). Capturing marketing information to fuel growth. *Journal of Marketing*, *85*(1), 163-183.
Farboodi, M., & Veldkamp, L. (2023). Data and markets. *Annual Review of Economics*, *15*(1), 23-40.
Ferracuti, E., Koo, M., Lee, M., & Stubben, S. (2024). Acquisition of customer information and corporate decision making. *Working paper.*
Franke, N., Schreier, M., & Kaiser, U. (2010). The "I designed it myself" effect in mass customization. *Management science*, *56*(1), 125-140.
Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. *Working paper*.
Han, P., Jiang, W., & Mei, D. (2024). Mapping US–China Technology Decoupling: Policies, Innovation, and Firm Performance. *Management Science*, *70*(12), 8386-8413.
He, J., & Tian, X. (2018). Finance and corporate innovation: A survey. *Asia-Pacific Journal of Financial Studies*, *47*(2), 165-212.
He, J., & Tian, X. (2020). Institutions and innovation. *Annual Review of Financial Economics*, *12*(1), 377-398.
He, Q., Leung, H., Qiu, B., & Zhou, Z. (2024). The effect of social media on corporate innovation: Evidence from seeking alpha coverage. *Management Science*.
Hirshleifer, D., Hsu, P. H., & Li, D. (2013). Innovative efficiency and stock returns. *Journal of Financial Economics*, *107*(3), 632-654.
Hoelzemann, J., Manso, G., Nagaraj, A., & Tranchero, M. (2024). The streetlight effect in data-driven exploration. *Working paper.*
Hsu, P. H., Lee, K., Moon, S. K., & Oh, S. (2024). Information disclosure and peer innovation: Evidence from mandatory reporting of clinical trials. *Journal of Financial and Quantitative Analysis* forthcoming.
Keeley, L., Walters, H., Pikkel, R., & Quinn, B. (2013). *Ten types of innovation: The discipline of building breakthroughs*. John Wiley & Sons.
Kong, D., Lin, C., Wei, L., & Zhang, J. (2022). Information accessibility and corporate innovation. *Management Science*, *68*(11), 7837-7860.
Kopalle, P. K., Kumar, V., & Subramaniam, M. (2020). How legacy firms can embrace the digital ecosystem via digital customer orientation. *Journal of the Academy of Marketing Science*, *48*, 114-131.
Lehrer, C., Wieneke, A., Vom Brocke, J. A. N., Jung, R., & Seidel, S. (2018). How big data analytics enables service innovation: materiality, affordance, and the individualization of service. *Journal of Management Information Systems*, *35*(2), 424-460.
Liszewski, A. (2025). *Apple says it'll use Apple Maps Look Around photos to train AI.* https://www.theverge.com/news/635316/apple-maps-look-around-photos-apple-intelligence-training.
Liu, G., Liu, H., Liu, Y., Yang, J., & Zhang, Y. (2024). Personal income tax and corporate innovation: The key role of inventors' financial incentives. *Journal of Banking & Finance*, *164*, 107203.
Merchant, B. (2023). *These apps and websites use your data to train AI. You're probably using one right now.* https://www.latimes.com/business/technology/story/2023-08-16/column-its-not-just-zoom-how-websites-and-apps-harvest-your-data-to-build-ai.
Nambisan, S. (2002). Designing virtual customer environments for new product development:

Toward a theory. *Academy of Management Review*, *27*(3), 392-413.
Saldanha, T. J., Mithas, S., & Krishnan, M. S. (2017). Leveraging customer involvement for fueling innovation. *MIS Quarterly*, *41*(1), 267-286.
Schweisfurth, T. G. (2017). Comparing internal and external lead users as sources of innovation. *Research Policy*, *46*(1), 238-248.
Stocchi, L., Pourazad, N., Michaelidou, N., Tanusondjaja, A., & Harrigan, P. (2022). Marketing research on Mobile apps: Past, present and future. *Journal of the Academy of Marketing Science*, 1-31.
Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175-199.
Tian, X., & Wang, T. Y. (2014). Tolerance for failure and corporate innovation. *The Review of Financial Studies*, *27*(1), 211-255.
Veldkamp, L. (2023). Valuing data as an asset. *Review of Finance*, *27*(5), 1545-1562.
Von Hippel, E. (2005). Democratizing innovation: The evolving phenomenon of user innovation. *Journal für Betriebswirtschaft*, *55*, 63-78.
Wu, X. (2023). Mobile app, firm risk, and growth. *Working paper*.
Xu, N., Chan, K. C., Xie, R., Chen, Q., & Agarwal, S. (2024). The impact of family-based human capital on corporate innovation: Evidence from sibling-chairpersons in China. *Management Science*, *70*(10), 7062-7089.
Yoo, Y., Boland Jr, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science*, *23*(5), 1398-1408.

Appendix 1 Examples of Using Mobile App Data in Patent Development

Example 1: Zhuhai Gree Electric Appliances Co., Ltd. (Patent Application Publication No. CN107504656A)

A Self-Learning Control Method and System for Air Conditioners

This invention discloses a self-learning control method and system for air conditioners. The system includes an information collection module and an autonomous control module. The information collection module gathers data on users' air conditioner usage, such as turning it on or off via a remote control or app, and transmits this data to the autonomous control module. The autonomous control module analyzes the collected data using a self-learning mechanism to identify the user's power-on and power-off patterns. Based on this analysis, it determines the air conditioner's operation schedule and automatically controls its power cycles. By learning from user behavior, this invention predicts when the air conditioner should be turned off, enabling intelligent and automated control.

Example 2: Haier Smart Home Co., Ltd. (Patent Application Publication No. CN115076764A)

A Heating Equipment Control Method and Heating Equipment

This invention discloses a heating equipment control method and corresponding heating equipment. The proposed control method includes the following steps: S1. Monitoring the app on the user’s mobile device to obtain data on mobile data consumption and/or consumption rate during the monitoring period. S2. Adjusting the indoor temperature based on the collected

data—if the data consumption is below a specified threshold and/or the consumption rate is lower than a designated rate, the system adjusts the indoor temperature to a predefined level.

By detecting app usage, this invention determines the user's living state and adjusts the heating equipment accordingly. It automatically regulates indoor temperature to align with the user's habits without requiring manual intervention, making the process more intelligent and convenient. Additionally, by analyzing the user’s activity patterns and the indoor-outdoor temperature difference, the system dynamically adjusts the heating power to ensure a smoother temperature transition, enhancing comfort and overall user experience.

Example 3: SAIC Motor Corporation Limited (Patent Application Publication No. CN112884515A)

User Churn Prediction Method, Device, and Computer Storage Medium

This application provides a user churn prediction method, device, and computer storage medium. The method involves obtaining historical data generated by a car-sharing rental app for the user to be predicted, extracting target features from the historical data, and inputting these features into a preset model to calculate the user’s churn probability. Since the target features accurately and comprehensively reflect the car-sharing rental scenario, the prediction model ensures high accuracy in churn probability estimation. Additionally, the model processes data at a high speed, enhancing the efficiency of user churn prediction while maintaining accuracy. This invention holds significant practical value for car-sharing rental service.

Appendix 2 App Examples and the Data They Collect

Total cumulative downloads are based on information as of January 2025.

Corporate name	App	App release date	Category	App main functions	App distribution regions	App download platforms	Total cumulative downloads(million)	Scope of collected data
37 Interactive Entertainment (002555.SZ)	Douluo Dalu	Oct 16, 2018	Gaming	Gaming	China	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	281	Transaction information (transaction records, order information); location data (network location); identifiers (user identifier, device identifier); user content (game data); app information (basic app information, app running logs); device information (installed apps list, hardware/software parameters, system settings, network information).
Haier Smart Home (600690.SH)	Haier Smart Home	June 9, 2015	Home equipment management	Connection, control, and information inquiry of electrical appliances; purchase, delivery, installation, maintenance, and repair of electrical appliances; membership benefits.	Globe	Apple App Store; Google Play Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	418	Contact information (phone number, address); health and fitness information (exercise data); transaction information (transaction records, order details, delivery information); location data (GPS location, network location); identifiers (user identifier, device identifier); basic information (account details); user content (images or videos, audio, text information, search terms, social interactions, customer support); app information (basic app information, app running logs); device information (hardware/software parameters, system settings, network information).
Changan Automobile (000625.SZ)	Changan Automobile	January 17, 2017	Automobile Management	Vehicle control, after-sales service, online store (marketplace), and car owners' community interaction.	Globe	Apple App Store; Google Play Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	49	Contact information (email address, phone number, contact list, address, social media accounts); financial information (bank account details); transaction information (transaction records, order information, order details, delivery information); location information (GPS location, network location); identifiers (user identifier, device identifier, identity information); basic information (name, gender, identity card, account information, account details); user content (images or videos, audio, text information, search queries, social interactions, customer support); app

							information (browsing history, favorites, basic app information, app operation logs, other usage data); vehicle-related information.
Shantui Construction Machinery (000680.SZ)	Shantui E-Path	August 12, 2019	Specialized Equipment Management	Initiate repair and maintenance requests; customers submit product improvement suggestions to the product manager, who determines whether to adopt the suggestions. For suggestions not adopted, a response is provided with the reason, and for adopted suggestions, the design improvement process is initiated; equipment management; information exchange.	Globe	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	0.3	Location data (GPS location)
Yonghui Superstores (601933.SH)	Yonghui Life	December 22, 2015	E-commerce	Shopping and payment	Globe	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	787	Contact information (phone number, address); transaction information (transaction records, order information); location information (GPS location, network location); identifiers (user identifier, device identifier, identity information); basic information (name, gender, date of birth, account information); user content (search terms); app information (browsing history, favorites, basic app information, app running logs, other usage data); device information (hardware/software parameters, system settings, network information).
Qingdao TGOOD Electric (300001.SZ)	Telepower	December 21, 2014	Life Services (Charging)	Query and locate charging stations with navigation; scan charging terminals; monitor vehicle battery levels, mileage, and	Globe	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360	59	Contact information (email address, phone number, address); transaction information (order details); location information (GPS location); identifiers (user identifier, device identifier, identity information); basic information (name, gender, age, date of birth, account information); user content (images or videos, text

				power consumption; track vehicle violation status.		App Store		messages, search terms); app information (browsing history, basic app information, app running logs); device information (installed app list, hardware/software parameters, system settings, network information).
China United Network Communications (600050.SH)	China Unicom	March 15, 2011	Telecommunication	Account inquiry; service handling; risk call interception; network security monitoring.	China	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	6062	Contact information (email address, phone number, contact list, address); transaction information (transaction records, order information, order details, delivery information); location information (GPS location, network location); identifiers (user identifier, device identifier, identity information); basic information (name, account information, account details); user content (images or videos, audio, text information, search queries, social interactions, customer support); app information (browsing history, favorites, basic app information, app operation logs, other usage data); device information (installed apps list, hardware/software parameters/system settings, network information, other device information).
Guangdong Lexin Medical Electronics (300562.SZ)	Leksun Health	September 18, 2015	Health and Medical Care	Health data collection; exercise programs; dietary guidance; health check-up guidance; online store.	China	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	93	Contact information (phone number); health and fitness information (exercise data).
Shenzhen MTC (002429.SZ)	Fengxing Video	September 18, 2015	Video	Playback of movies, TV dramas, animations, variety shows, sports programs, and more.	China	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	491	Location information (network location); identifiers (user identifier, device identifier); basic information (account information); user content (text information, search terms, social interactions, customer support); app information (browsing history, favorite records, basic app information, app running logs, other app usage data); device information (installed apps list, hardware/software parameters, system settings, network information, other device information).
Beijing Dabeinong Technology (002385.SZ)	Zhinongtong	March 18, 2015	Planting and breeding	Online consultation with veterinarians; veterinary medicine marketplace; disease prevention and treatment techniques sharing and exchange;	Globe	Apple App Store; Huawei App Store; Tencent App Store; Xiaomi App Store; 360 App Store	11	Contact information (email address, phone number, contact list, address, social media accounts); financial information (bank account details); transaction information (transaction records, order details, delivery information); location information (GPS location, network location); sensitive information (biometric data); identifiers (user identifiers, device identifiers,

		livestock news push notifications; online transactions.	identity information); basic information (name, gender, age, date of birth, employment details, account information); user content (images or videos, audio, text messages, search terms, social interactions); application information (browsing history, favorites, basic app information, app running logs); device information (installed apps list, hardware/software parameters, system settings, network information, other device details).

Appendix 3 Sample Selection Process

This table presents the sample selection. If a company has an App during the sample period, the corresponding observations are defined as the treatment group. Conversely, if a company has not launched an App by the end of the sample period, the corresponding observations are defined as the control group. The table reports the number of companies in the treatment and control groups, with the number of firm-year observations shown in parentheses.

	Treatment group	Control group
Chinese A-share listed firms between 2007 and 2021	760(7,033)	4010(35,142)
After dropping financial and software and related services industry firms	527(5,028)	3867(33,755)
Retain firms in industry with at least 5% firms that have launched mobile Apps before the end of the sample period	500(4,699)	3255(26,901)
Firms in treatment group are required to have at least one year of data both before and after App release	343(4,119)	3255(26,901)
Retain firms with at least one eventually granted invention patent filed and cited during the sample period	311(3,754)	2576(22,787)
Exclude firm-years with missing values of the variables and firms with special treatment (ST or *ST)	308(3,643)	2274(21,018)

Appendix 4 Variable Definition

Variable	Definition	Source
Pat	The total number of invention patent applications filed by a firm in a year that are eventually granted. Patents data are adjusted for truncation bias using the application-grant lag distribution.	CNRDS
Cite	The number of forward citations to all eventually granted invention patents filed by a firm in a year, excluding self-citations. Citations are adjusted for the truncation problem using the "fixed-effect" approach.	CNRDS
LnPat	Ln(1+*Pat*)	CNRDS
LnCite	Ln(1+*Cite*)	CNRDS
DigitalPat_shr	*DigitalPat*/(1+*Pat*). *DigitalPat* is the number of patent applications related to digital innovation filed by a firm in a year.	CNRDS
DigitalCite_shr	*DigitalCite*/(1+*Cite*). *DigitalCite* is the number of citations count of digital patents filed by a firm in a year, excluding self-citations.	CNRDS
LnPatR&D	The natural logarithm of one plus the ratio of the number of successful invention patents filed to R&D expenditure in millions RMB.	CNRDS
Approval	The natural logarithm of one plus the ratio of granted patents over total patent applications.	CNRDS
UserPat_shr	*UserPat*/(1+*Pat*). *UserPat* is the number of user-relevant patent applications filed by a firm in a year.	CNRDS
UserCite_shr	*UserCite*/(1+*Cite*). *UserCite* is the number of citations count of user-relevant patents filed by a firm in a year, excluding self-citations.	CNRDS
Applaunch	An indicator variable equals 1 if the firm has launched at least one mobile app in a year, and 0 otherwise.	Qimai
Applaunch_More	A dummy variable equals one if the firm has released an app and the app collects more user data, and zero otherwise. We use three distinct criteria to identify whether the app collects more user data: (1) the app download volume is above the median, (2) the app reviewer volume is above the median, and (3) the app directly collects user-linked data, according to its data collection policy.	Qimai; App Store
Applaunch_Less	A dummy variable equals one if the firm has released an app and the app collects less user data, and zero otherwise. Similarly, we use three distinct criteria to identify whether the app collects less user data: (1) the app download volume is below the median, (2) the app reviewer volume is below the median, and (3) the app does not directly collect user-linked data, according to its data collection policy.	Qimai; App Store

Applaunch_Comp	A dummy variable equals one if the firm has released an app and the app has not been criticized by regulatory authorities, and zero otherwise.	MIIT website
Applaunch_NonComp	A dummy variable equals one if the firm has released an app and the app has been criticized by regulatory authorities, and zero otherwise.	MIIT website
Size	The natural logarithm of total assets at the end of the year. Total assets are measured in million RMB.	CSMAR
BM	The ratio of the book value of equity to the market value of equity of a firm in a year.	CSMAR
Leverage	The ratio of total debt to total assets of a firm in a year.	CSMAR
Cash	Cash holdings scaled by total assets of a firm in a year.	CSMAR
ROA	The ratio of net income to total assets of a firm in a year.	CSMAR
Age	The natural logarithm of firm listing age.	CSMAR
InstOwn	Institutional ownership of a firm in a year.	CSMAR
OwnCon	The ratio of shareholdings of the largest shareholder.	CSMAR
PPE	Property, plant, and equipment divided by total assets.	CSMAR
R&D	R&D expenditure scaled by total assets of a firm in a year, multiply by 100. We denote a missing R&D observation as zero.	CSMAR
HHI	The Herfindahl-Hirschman index based on annual sales using CSRC industry codes.	CSMAR
HHI^2	The squared term of Herfindahl–Hirschman index.	CSMAR

Appendix 5 Dictionary

Digital Relevant

Original English Keywords:

AI | H5 | 3D | 5G | IT | DBA | SQL | NLP | APP | SDK | Web | GIS | IOS | Net | Java | ABAP | Linux | Hadoop | Python | Android | Windows | Fintech | Digital

Chinese Keywords (translated):

C language | C software | NFC payment | EB-level storage | Algorithm | Programming | Cloud | Cloud network | Virtual | Smart | Intelligent | Android | Digitalization | Unmanned | Networking | Informatization | Automation | Visualization | Integration | Unification | Standardization | Modularization | Electronization | Cloud application | Internet of Things (IoT) | Blockchain | Internet | Cloud platform | Cloud service | Cloud computing | Cloud technology | Virtualization | Robot | Cloud security | Cloud ecosystem | Cloud system | Cloud product | Graph computing | Stream computing | Server | Mini program | Mobile end | Neural network | Smart city | Smart transportation | Image recognition | Image understanding | Virtual reality | Biometric recognition | Pattern recognition | Network security | Semantic understanding | Facial recognition | User profiling | Customer profiling | Speech recognition | Human-computer interaction | Semantic analysis | Triple play (Three-network convergence) | Information security | Network security | Mobile payment | Mobile healthcare | Equity crowdfunding | Autonomous driving | Self-driving | Mobile payment | Cutting-edge technology | Machine learning | Deep learning | Semantic search | Identity verification | In-memory computing | Autonomous driving | Text mining | Augmented reality | Mixed reality | Brain-inspired computing | Green computing | Converged architecture | Billion-level concurrency | Cognitive computing | Digital currency | E-commerce | Smart agriculture | Digital marketing | Unmanned retail | Digital finance | FinTech | Quantitative finance | Open banking | Information technology | Information system | Software development | System development | Edge computing | Natural language | Algorithm engineer | Design engineer | R&D engineer | Development engineer | Operations engineer | Software engineer | Software developer | Software architect | Test engineer | Electronic engineer | Distributed computing | Third-party payment | Cyber-physical system | Secure multi-party computation | Differential privacy technology | Natural language processing | Decision support system | Data

User Relevant

Chinese Keywords (translated):

User | Consumer | Customer | Client | Personal | Private | Identity | Interaction | Engagement | Participation | Community | Communication | Attitude | Behavior | Habit | Interface | Usability | Compatibility | Practicality | Accessibility | Comprehensibility | Ease of use | Operation guide | Operation process | Customization | Personalization | Tailor-made | Individualized | Specialized | Suitable | Adaptive | Experience | Convenience | Trust | Loyalty | Retention | Brand | Real-time | Feedback | Response | Needs | Preferences | Likes | Affinity | Desire | Expectation | Anticipation | Satisfaction | Contentment | Service | Service upgrade | Solution | Product design | Scalability | Terminal | Consumption | Purchase | Browsing | Visit | Click | Transaction | Payment | Mobile | App | Mobile software | Application software | Program | Mini program | Intelligence | Data security | Intelligent recommendation | Immersive | Speech recognition | Virtual reality | Recommendation system | Multi-platform | Visualization | Automation

Table 1 Apps Distribution

Panel A shows the sample distribution of firms by year. Panel B shows the sample distribution by industry. Industry classifications are from the CSRC industry classification scheme with two-digit industry codes for manufacturers and one-digit industry codes for other sectors.

Panel A Apps Distribution by Year

	Initial sample		Final sample	
	#	%	#	%
2009	2	0.26		
2010	17	2.24	2	0.65
2011	31	4.08	9	2.92
2012	66	8.68	11	3.57
2013	69	9.08	22	7.14
2014	76	10.00	25	8.12
2015	120	15.79	58	18.83
2016	103	13.55	40	12.99
2017	94	12.37	41	13.31
2018	60	7.89	33	10.71
2019	44	5.79	25	8.12
2020	50	6.58	23	7.47
2021	28	3.68	19	6.17
Total	760	100.00	308	100.00

Panel B Apps Distribution by Industry

	Initial sample		Final sample	
Industry	#	%	#	%
Agriculture	2	0.26		
Mining	2	0.26		
Food and Beverage	26	3.42	20	6.49
Textile, Apparel, Fur, and Leather	9	1.18	4	1.30
Furniture Manufacturing	9	1.18	5	1.62
Paper and Printing	13	1.71	8	2.60
Petroleum, Chemical, Plastics, and Rubber	19	2.50		
Electronics	77	10.13	52	16.88
Metal and Non-Metal	20	2.63	17	5.52
Machinery, Equipment, and Instruments	129	16.97	92	29.87
Medicine and Biological Products	19	2.50	11	3.57
Other Manufacturing	3	0.39	2	0.65
Energy and Water	6	0.79		
Construction	10	1.32	8	2.60
Transportation	27	3.55	15	4.87
IT and Computing	171	22.50	12	3.90
Wholesale and Retail Trade	44	5.79	21	6.82
Finance	91	11.97		
Real Estate	21	2.76	16	5.19
Public Utilities	38	5.00	15	4.87
Entertainment	23	3.03	9	2.92
Conglomerates	1	0.13	1	0.32
Total	760	100.00	308	100.00

Table 2 Summary Statistics

The definitions of variables are presented in Appendix 4.

Variable	Obs.	Mean	S.D.	Q1	Median	Q3
Pat_{t+1}	24,661	23.431	85.970	0.000	3.475	13.900
$Cite_{t+1}$	24,661	11.177	36.053	0.000	1.590	6.876
Applaunch	24,661	0.067	0.251	0.000	0.000	0.000
Size	24,661	8.346	1.301	7.403	8.148	9.073
BM	24,661	0.445	0.324	0.221	0.360	0.564
Leverage	24,661	0.431	0.205	0.267	0.425	0.587
Cash	24,661	0.190	0.133	0.097	0.153	0.243
ROA	24,661	0.039	0.060	0.015	0.038	0.068
Age	24,661	2.030	0.897	1.386	2.197	2.773
InstOwn	24,661	0.449	0.246	0.253	0.470	0.648
OwnCon	24,661	34.673	14.714	23.180	32.770	44.610
PPE	24,661	0.200	0.142	0.091	0.172	0.277
R&D	24,661	1.827	1.724	0.337	1.566	2.686
HHI	24,661	0.095	0.098	0.040	0.066	0.113
HHI^2	24,661	0.019	0.055	0.002	0.004	0.013

Table 3 Baseline Results

This table presents the results of DiD analysis using the Poisson model to examine the effect of releasing a mobile App on corporate innovation. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively.

	(1)	(2)
	Pat_{t+1}	$Cite_{t+1}$
Applaunch	0.154**	0.153***
	(0.071)	(0.058)
Size	0.363***	0.151***
	(0.045)	(0.052)
BM	0.235***	0.085
	(0.066)	(0.057)
Leverage	-0.130	0.193
	(0.191)	(0.203)
Cash	-0.084	0.059
	(0.226)	(0.180)
ROA	2.014***	2.310***
	(0.312)	(0.396)
Age	-0.073	0.012
	(0.056)	(0.056)
InstOwn	0.735***	0.382**
	(0.178)	(0.186)
OwnCon	-0.003	-0.004
	(0.003)	(0.003)
PPE	-0.626**	-0.454*
	(0.281)	(0.264)
R&D	0.068***	0.043***
	(0.016)	(0.015)
HHI	0.044	0.753
	(0.842)	(0.789)
HHI^2	-0.884	-1.775*
	(1.072)	(1.049)
Constant	0.647	1.926***
	(0.501)	(0.531)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	24,661	24,661
Pseudo R^2	0.872	0.820

Table 4 Dynamic Test

This table presents the results of the dynamic test. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively.

	(1)	(2)
	Pat_{t+1}	$Cite_{t+1}$
Applaunch (-3^{+})	-0.049	-0.048
	(0.081)	(0.082)
Applaunch (-2)	-0.050	-0.048
	(0.037)	(0.043)
Applaunch (0)	0.070	0.065
	(0.046)	(0.043)
Applaunch (+1)	0.091*	0.137**
	(0.053)	(0.056)
Applaunch ($+2^{+}$)	0.145**	0.144**
	(0.067)	(0.070)
Constant	0.646	1.929***
	(0.502)	(0.531)
Controls	Yes	Yes
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	24,661	24,661
Pseudo R^2	0.872	0.820

Table 5 Controlling for Other Interactive Fixed Effects

This table presents the results after controlling for interactive fixed effects. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Pat_{t+1}	$Cite_{t+1}$	Pat_{t+1}	$Cite_{t+1}$	Pat_{t+1}	$Cite_{t+1}$
Applaunch	0.130*	0.154***	0.173***	0.159***	0.151**	0.167***
	(0.069)	(0.058)	(0.061)	(0.050)	(0.060)	(0.049)
Constant	-0.105	1.586***	0.499	2.148***	-0.163	1.795***
	(0.521)	(0.536)	(0.465)	(0.509)	(0.473)	(0.478)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	No	No	Yes	Yes
Province-Year FE	No	No	Yes	Yes	Yes	Yes
Obs.	24,654	24,654	24,651	24,635	24,644	24,628
Pseudo R^2	0.878	0.826	0.879	0.828	0.884	0.834

Table 6 Stacked-cohort DiD Tests Based on Matched Sample

This table presents the results using the stacked-cohort DiD approach. Panel A shows the univariate comparisons of matching variables after PSM. Panel B shows the results of stacked DiD with matched sample. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively.

Panel A The Differences between Treatment and Control Group after Matching

Variable	Mean of treatment group (Cohort-firm=296)	Mean of control group (Cohort-firm=872)	Difference of mean	p-value
$Size_{average(t-2\ to\ t-1)}$	8.846	8.715	0.131	0.170
$BM_{average(t-2\ to\ t-1)}$	0.458	0.463	-0.005	0.811
$Leverage_{average(t-2\ to\ t-1)}$	0.416	0.413	0.003	0.830
$Cash_{average(t-2\ to\ t-1)}$	0.232	0.219	0.013	0.168
$ROA_{average(t-2\ to\ t-1)}$	0.0590	0.0560	0.004	0.198
$Age_{average(t-2\ to\ t-1)}$	1.854	1.901	-0.047	0.479
$InstOwn_{average(t-2\ to\ t-1)}$	0.502	0.505	-0.004	0.831
$OwnCon_{average(t-2\ to\ t-1)}$	37.94	37.76	0.174	0.867
$PPE_{average(t-2\ to\ t-1)}$	0.181	0.182	-0.001	0.898
$R\&D_{average(t-2\ to\ t-1)}$	2.007	1.961	0.046	0.700
$LnPat_{average(t-2\ to\ t-1)}$	1.919	1.868	0.051	0.625

Panel B Regression Results Based on Matched Sample

	(1) Pat_{t+1}	(2) $Cite_{t+1}$
Applaunch	0.162**	0.181***
	(0.065)	(0.063)
Constant	1.882***	2.961***
	(0.640)	(0.809)
Controls	Yes	Yes
Cohort-firm FE	Yes	Yes
Cohort-year FE	Yes	Yes
Obs.	8,617	8,583
Pseudo R^2	0.908	0.882

Table 7 Heterogeneity-robust DiD Estimations

This table presents the results of heterogeneity-robust DiD estimations. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-sided tests, respectively.

	Sun and Abraham (2021)		Borusyak et al. (2024)	
	$LnPat_{t+1}$	$LnCite_{t+1}$	$LnPat_{t+1}$	$LnCite_{t+1}$
Applaunch	0.257***	0.179***	0.251***	0.160***
	(0.058)	(0.057)	(0.057)	(0.056)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Obs.	24,661	24,661	24,661	24,661
Adj. R^2	0.705	0.686	-	-

Table 8 Innovation Efficiency

This table presents the results of innovation efficiency. We measure innovation efficiency based on the number of successful patent applications per R&D expenses and the success rate of patent application. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively.

	(1) Successful patents per R&D expense $LnPatR\&D_{t+1}$	(2) Success rate of patent applications $Approval_{t+1}$
Applaunch	0.027**	0.028**
	(0.014)	(0.014)
Size	-0.029***	0.051***
	(0.007)	(0.007)
BM	-0.016	-0.004
	(0.013)	(0.013)
Leverage	-0.003	-0.088***
	(0.026)	(0.028)
Cash	-0.022	0.017
	(0.024)	(0.027)
ROA	0.098***	0.167***
	(0.035)	(0.050)
Age	0.027***	-0.026***
	(0.008)	(0.008)
InstOwn	0.030	0.033
	(0.028)	(0.029)
OwnCon	-0.000	-0.000
	(0.000)	(0.000)
PPE	-0.105***	0.007
	(0.039)	(0.037)
R&D	-	0.012***
	-	(0.002)
HHI	0.073	0.253**
	(0.129)	(0.117)
HHI^2	-0.124	-0.535***
	(0.214)	(0.178)
Constant	0.360***	0.046
	(0.059)	(0.059)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	21,110	22,207
Adj. R^2	0.354	0.358

Table 9 Data-driven Innovation

This table presents the results of data-driven innovation. Panel A shows the heterogeneous tests based on data volume. Panel B shows the effect of releasing App on corporate digital innovation. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively. After excluding apps with missing information on download volume and reviewer volume, 252 apps remain in the regression. This results in a reduced number of observations (i.e., 24,150) in Columns (1) to (4) of Panel A.

Panel A Heterogeneous Tests Based on Data Scale

	(1)	(2)	(3)	(4)	(5)	(6)
	App download volume		App reviewer volume		Collect user-linked data	
	Pat_{t+1}	$Cite_{t+1}$	Pat_{t+1}	$Cite_{t+1}$	Pat_{t+1}	$Cite_{t+1}$
Applaunch_More	0.207**	0.179**	0.236**	0.222***	0.195**	0.190***
	(0.092)	(0.079)	(0.099)	(0.083)	(0.080)	(0.068)
Applaunch_Less	-0.088	0.083	-0.077	0.018	-0.033	-0.006
	(0.111)	(0.092)	(0.118)	(0.090)	(0.134)	(0.089)
Size	0.375***	0.153***	0.372***	0.153***	0.360***	0.149***
	(0.045)	(0.052)	(0.044)	(0.052)	(0.045)	(0.051)
BM	0.218***	0.076	0.211***	0.074	0.236***	0.084
	(0.065)	(0.059)	(0.066)	(0.059)	(0.065)	(0.057)
Leverage	-0.162	0.185	-0.178	0.176	-0.126	0.196
	(0.189)	(0.205)	(0.189)	(0.203)	(0.190)	(0.202)
Cash	-0.056	0.061	-0.044	0.060	-0.081	0.059
	(0.228)	(0.185)	(0.229)	(0.186)	(0.225)	(0.180)
ROA	1.953***	2.286***	1.931***	2.275***	2.009***	2.309***
	(0.312)	(0.398)	(0.308)	(0.399)	(0.309)	(0.397)
Age	-0.071	0.004	-0.070	0.004	-0.063	0.021
	(0.057)	(0.059)	(0.056)	(0.059)	(0.055)	(0.055)
InstOwn	0.739***	0.398**	0.723***	0.386**	0.735***	0.386**
	(0.179)	(0.187)	(0.178)	(0.186)	(0.177)	(0.185)
OwnCon	-0.003	-0.004	-0.003	-0.004	-0.003	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
PPE	-0.588**	-0.490*	-0.590**	-0.492*	-0.627**	-0.462*
	(0.286)	(0.268)	(0.285)	(0.268)	(0.280)	(0.264)
R&D	0.066***	0.042***	0.064***	0.041***	0.067***	0.042***
	(0.016)	(0.015)	(0.016)	(0.015)	(0.016)	(0.015)
HHI	-0.092	0.717	-0.147	0.746	0.117	0.810
	(0.839)	(0.804)	(0.849)	(0.809)	(0.838)	(0.791)
HHI^2	-0.850	-1.813*	-0.773	-1.880*	-1.032	-1.918*
	(1.089)	(1.073)	(1.104)	(1.085)	(1.085)	(1.064)
Constant	0.540	1.923***	0.593	1.931***	0.639	1.919***
	(0.495)	(0.535)	(0.486)	(0.530)	(0.498)	(0.527)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	24,150	24,150	24,150	24,150	24,661	24,661
Pseudo R^2	0.870	0.816	0.870	0.816	0.872	0.820
P-value of coef. diff.	0.036	0.411	0.040	0.092	0.141	0.075

Panel B Effect on Digital Innovation

	(1)	(2)
	$DigitalPat_shr_{t+1}$	$DigitalCite_shr_{t+1}$
Applaunch	0.022***	0.020***
	(0.006)	(0.008)
Size	0.008***	0.008***
	(0.002)	(0.002)
BM	0.007	0.010**
	(0.004)	(0.005)
Leverage	0.008	0.008
	(0.009)	(0.010)
Cash	-0.004	0.003
	(0.011)	(0.012)
ROA	0.031*	0.048**
	(0.019)	(0.019)
Age	-0.003	-0.009***
	(0.003)	(0.003)
InstOwn	0.004	0.014
	(0.009)	(0.010)
OwnCon	-0.000	-0.000**
	(0.000)	(0.000)
PPE	0.002	0.008
	(0.011)	(0.013)
R&D	0.003***	0.003**
	(0.001)	(0.001)
HHI	-0.044	-0.025
	(0.038)	(0.041)
HHI^2	0.030	-0.015
	(0.067)	(0.067)
Constant	-0.009	-0.005
	(0.019)	(0.020)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	24,661	24,661
Adj. R^2	0.437	0.389

Table 10 The Regulatory Compliance of The App

This table presents the results of heterogeneous tests based on the regulatory compliance of the app. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively.

	(1)	(2)
	Pat_{t+1}	$Cite_{t+1}$
Applaunch_Comp	0.130*	0.129*
	(0.078)	(0.066)
Applaunch_NonComp	0.303*	0.252**
	(0.166)	(0.114)
Size	0.361***	0.150***
	(0.045)	(0.052)
BM	0.235***	0.085
	(0.066)	(0.057)
Leverage	-0.123	0.196
	(0.191)	(0.202)
Cash	-0.102	0.042
	(0.230)	(0.183)
ROA	2.019***	2.295***
	(0.311)	(0.393)
Age	-0.070	0.013
	(0.056)	(0.056)
InstOwn	0.735***	0.383**
	(0.178)	(0.186)
OwnCon	-0.003	-0.004
	(0.003)	(0.003)
PPE	-0.640**	-0.467*
	(0.282)	(0.265)
R&D	0.067***	0.042***
	(0.016)	(0.015)
HHI	0.047	0.758
	(0.847)	(0.791)
HHI^2	-0.884	-1.791*
	(1.087)	(1.054)
Constant	0.653	1.936***
	(0.501)	(0.530)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	24,661	24,661
Pseudo R^2	0.872	0.820
P-value of coeff. diff.	0.342	0.343

Table 11 User-centric Innovation

This table presents the results of user-centric innovation. The definitions of variables are presented in Appendix 4. Robust standard errors adjusted for firm clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels based on two-tailed tests, respectively.

	(1)	(2)
	UserPat_shr$_{t+1}$	*UserCite_shr*$_{t+1}$
Applaunch	0.023***	0.015*
	(0.009)	(0.009)
Size	0.010***	0.010***
	(0.003)	(0.003)
BM	0.022***	0.020***
	(0.006)	(0.006)
Leverage	-0.001	-0.003
	(0.013)	(0.013)
Cash	-0.017	-0.021
	(0.015)	(0.014)
ROA	0.059**	0.072***
	(0.025)	(0.023)
Age	-0.014***	-0.021***
	(0.005)	(0.004)
InstOwn	-0.001	0.010
	(0.013)	(0.013)
OwnCon	-0.000	-0.000
	(0.000)	(0.000)
PPE	-0.030*	-0.021
	(0.017)	(0.017)
R&D	0.006***	0.006***
	(0.002)	(0.002)
HHI	-0.070	-0.022
	(0.058)	(0.059)
HHI^2	0.044	-0.025
	(0.096)	(0.096)
Constant	0.043	0.038
	(0.028)	(0.027)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Obs.	24,661	24,661
Adj. R^2	0.406	0.384